REMARKS

In the patent application, claims 1, 3-48 are pending. In the office action, claims 1, 3-38 and 40-48 claims are rejected. Claim 39 has not been properly rejected in this office action.

Applicant has amended claims 1, 4-7, 11-15, 19, 22-27, 31 and 32 to change the wording in the claims.

No new matter has been introduced.

At section 5 of the office action, claims 1, 3-14, 19-38 and 40-48 are rejected under 35 U.S.C. 102(b) as being anticipated by *Gersho et al.* (U.S. Patent No. 6,311,154, hereafter referred to as *Gersho*).

In rejecting claim 1, the Examiner states that *Gersho* discloses segmenting the audio signal into a plurality of segments based on the audio characteristics of the audio signal. The Examiner considers segmenting too be the same as partitioning or classifying.

It is respectfully submitted that while partitioning is analogous to segmenting, classifying is not the same as segmenting. In plain English, segmenting is to divide into segments and classifying is to arrange the segments according to category. The Examiner also considers frames as being the same as segments. The Examiner also points to col. 4, lines 25-27 of *Gersho* to show that *Gersho* discloses segmenting the audio signal into a plurality of segments based on the audio characteristics of the audio signal.

Applicant respectfully disagrees.

At col.4, lines 23 –34, *Gersho* discloses a method for coding a speech signal with the following steps:

a. partitioning samples of a speech signal in frames;
b. deriving a residual signal for each frame;
c. classifying the speech signal in each frame into a plurality of classes;
d. identifying the location of at least one window in the frame by examining the residual signal for the frame;

e. encoding the excitation for the frame using one of a plurality of excitation coding techniques selected according to the class of frames; and

f. confining all or substantially all of non-zero excitation amplitudes to lie within the windows.

From the above-description, it is clear that *Gersho* already determines where to set the boundary of each segment when partitioning the speech samples without knowing the audio characteristics of the speech signal in the segments. *Gersho* partitions the speech sample into frames. After partitioning in step (a), *Gersho* classifies the speech signal into classes in step (c). In other words, in *Gersho*, the samples of speech signal are first partitioned into frames and each frame is then classified into one of a plurality of classes. Before classifying, it is impossible to partition the speech signal based on the classes. *Gersho* does not partition the speech signal in frames based on the classes as stated by the Examiner. The partitioning step in *Gersho* is carried out independently of the audio characteristics of the speech signal.

In contrast, according to the present invention, the speech signals are partitioned into segments based on the audio characteristics in the speech signals. How the speech signals are partitioned depends on the audio characteristics of the speech signal. Because the audio characteristics of the audio signal vary from sample to sample, the boundary of the segments cannot be pre-determined. As a result, a segment can be long or short; it can be 10 frames or 28 frames (see Figure 3). In *Gersho*, the length of each partitioned "segment" is the same.

Thus, *Gersho* does not disclose or even suggest segmenting the audio signal into a plurality of segments based on the audio characteristics of the audio signal.

In this office action, the Examiner points to the Abstract to show that *Gersho* does teach segmenting the audio signal into a plurality of segments based on the audio characteristics of the audio signal. In particular, the Examiner states that *Gersho* discloses that "the speech is partitioned into frames and sub-frames. Performance is enhanced by coding the important segments of the excitation more accurately" (Abstract).

The entire Abstract of *Gersho* is shown below:

A speech coder and a method for speech coding wherein the speech signal is represented by an excitation signal applied to a synthesis filter. The speech is partitioned into frames and subframes. A classifier identifies which of the several categories the speech frame belongs to, and a different coding method is applied to represent the excitation for each category. For some categories, one or more windows are identified for the frame where all or most of the excitation signal samples are assigned by a coding scheme. Performance is enhanced by coding the important segments of the excitation more accurately. The window locations are determined from a linear prediction residual by identifying peaks of the smoothed residual energy contour. The method adjusts the frame and subframe boundaries so that each window is located entirely within a modified subframe or frame. This eliminates the artificial restriction incurred when coding a frame or subframe in isolation, without regard for the local behavior of the speech signal across frame or subframe boundaries.

The Abstract clearly shows that classification is carried out by a classifier after the speech is partitioned into frames and subframes. After classification, frames belonging to a category are coded by a coding method that represents the excitation in those categories. In the frames belonging to certain categories, one or more windows are identified for the frame so that all or most of the excitation signal samples within the frame are assigned to a coding scheme. In order to improve the performance, the important segments of the excitation within the frame are coded more accurately.

For the improvement of the performance, *Gersho* also discloses the following steps after the frames of the speech signal are classified (col.4, lines 22-34):

- identifying the location of at least one window in the frame by examining the residual signal for the frame;
- encoding the excitation for the frame using one of a plurality of excitation coding techniques, selected according to the class of the frame; and, for at least one of the classes, and
- confining all or substantially all of non-zero excitation amplitudes to lie within the windows.

In order to carry out these steps, *Gersho* partitions each fixed frame into M equal or nearly equal length "basic" subframes (col.7, lines 18-26). Each of the basic subframes is associated with a search subframe. The search subframe is adapted such that it contains an integer number of windows (col.8, lines 13-23). Each of these windows is the actual time location of the active intervals of the excitation signal (col.7, lines 40-47). The location and duration of the window is adapted to suit the local characteristics of the speech (col.7, lines 57-60).

Gersho discloses that the windows are coded differently depending on the classification of the speech frames: strongly periodic, weakly periodic, erratic and unvoiced (col.9, lines 44-55). The coding methods can be tailored to each frame category as described at col. 10, line 10 to col. 11, line 61. In order to reduce the bit rate in coding, the windows are not coded independently. Rather, the excitation signal is coded for each search subframe. This is possible because there is considerable correlation between the excitation signal in different windows in the same subframe, especially when the speech segment is periodic (col. 9, lines 17-29). For example, the excitation signal in multiple windows located in the same subframe is constrained to have the same fixed codebook contribution (col. 14, lines 17-19). *Gersho* also describes how the window is positioned in the subframe and how the location of the window is identified so that a subframe or frame can be modified so that the window lies entirely within the modified subframe or frame (col.4, line 56 to col. 5, line 6).

In sum, in order to enhance the coding efficiency, *Gersho* discloses coding the excitation signal in the windows depending on the classification of the speech frames. *Gersho* also discloses dividing a fixed frame into a number of subframes for the purpose of locating the active periods (i.e., windows) of the excitation signal in the subframes. However, *Gersho* does not disclose or suggest segmenting each fixed frame into a plurality of subframes based on the audio characteristics of the audio signal in the fixed frame. *Gersho* only discloses coding the excitation in the subframes depending on the audio characteristics of the fixed frame. *Gersho* does not disclose or suggest segmenting the speech signal into a plurality of fixed frames based on the audio characteristics of the speech signal. *Gersho* only discloses classifying the speech signal in each of the fixed frames into different classes using two classifiers (col.4, lines 51-55).

Furthermore, *Gersho* is irrelevant to the present invention, because the coding method in *Gersho* is completely different from the coding method of the present invention as claimed. The claimed invention is concerned with a parametric-type encoding method, whereas *Gersho* is concerned with a CELP-type encoding method.

In the parametric-type encoding method, a parametric speech production model is used to obtain a set of parameters from the audio signal so as to produce a further audio signal in the decoder based on the parameters. The parametric-type encoding and decoding method, as disclosed in the specification does not rely on the waveform of the speech signal segments. In fact, due to the loss of the synchrony between the coder input and output signal, waveform matching is not carried out.

A CELP coder is an example of an Analysis-by-Synthesis (AbS) coder (see col.1, line 54 to col.2, line 1). As known in the art, a CELP coder performs waveform matching on the coder output using code excitation candidates and selecting the one minimizing given error criteria. As disclosed in *Gersho*, the CELP coder relies on the residual and excitation models. *Gersho*'s coder is not a parametric coder as disclosed in the present invention.

For the above reasons, claims 1, 19, 22, 26, 27, 31 and 32 are clearly distinguishable over *Gersho*.

As for claims 3-14, 20, 21, 23-25, 28-30 and 33-48, they are dependent from claims 1, 19, 22, 26, 27 and 31 and recite features not recited in claims 1, 19, 22, 26, 27 and 31. For reasons regarding claims 1, 19, 22, 26, 27 and 31 above, it is respectfully submitted that claims 3-14, 20, 21, 23-25, 28-30 and 33-48 are distinguishable over the cited *Gersho* reference.

At section 7, claims 15-18 are rejected under 35 U.S.C. 103(a) as being unpatentable over *Gersho*, in view of *Gersho IEEE-96*. The Examiner cites *Gersho IEEE-96* for disclosing the limitations in claims 15-18.

It is respectfully submitted that claims 15-18 are dependent from claims 1 and recite features not recited in claim 1. For reasons regarding claim 1 above, claims 15-18 are also distinguishable over the cited *Gersho* and *Gersho IEEE-96* references.

CONCLUSION

Claims 1, 3-48 are allowable. Early allowance of these pending claims is earnestly solicited.

Respectfully submitted,

Ken Lao

Kenneth Q. Lao

Attorney for the Applicant
Registration No. 40,061

WARE, FRESSOLA, VAN DER SLUYS
& ADOLPHSON LLP
Bradford Green, Building Five
755 Main Street, P.O. Box 224
Monroe, CT 06468
Telephone: (203) 261-1234
Facsimile: (203) 261-5676
USPTO Customer No. 004955